RSAC 2025 Conference

Many Voices. One Community.

SESSION ID: PDP-T01

RAG-NAROK: What Poorly-Built RAGs Can Do to Data Security

Akash Mukherjee

Head of Engineering
Realm Labs
https://linkedin.com/in/akashmuk

Saurabh Shintre, PhD

Founder and CEO
Realm Labs
https://linkedin.com/in/saurabh-shintre-3b47465/

Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

About us

Dr. Saurabh Shintre
CEO, Realm Labs

Akash Mukherjee
Head of Eng, Realm Labs

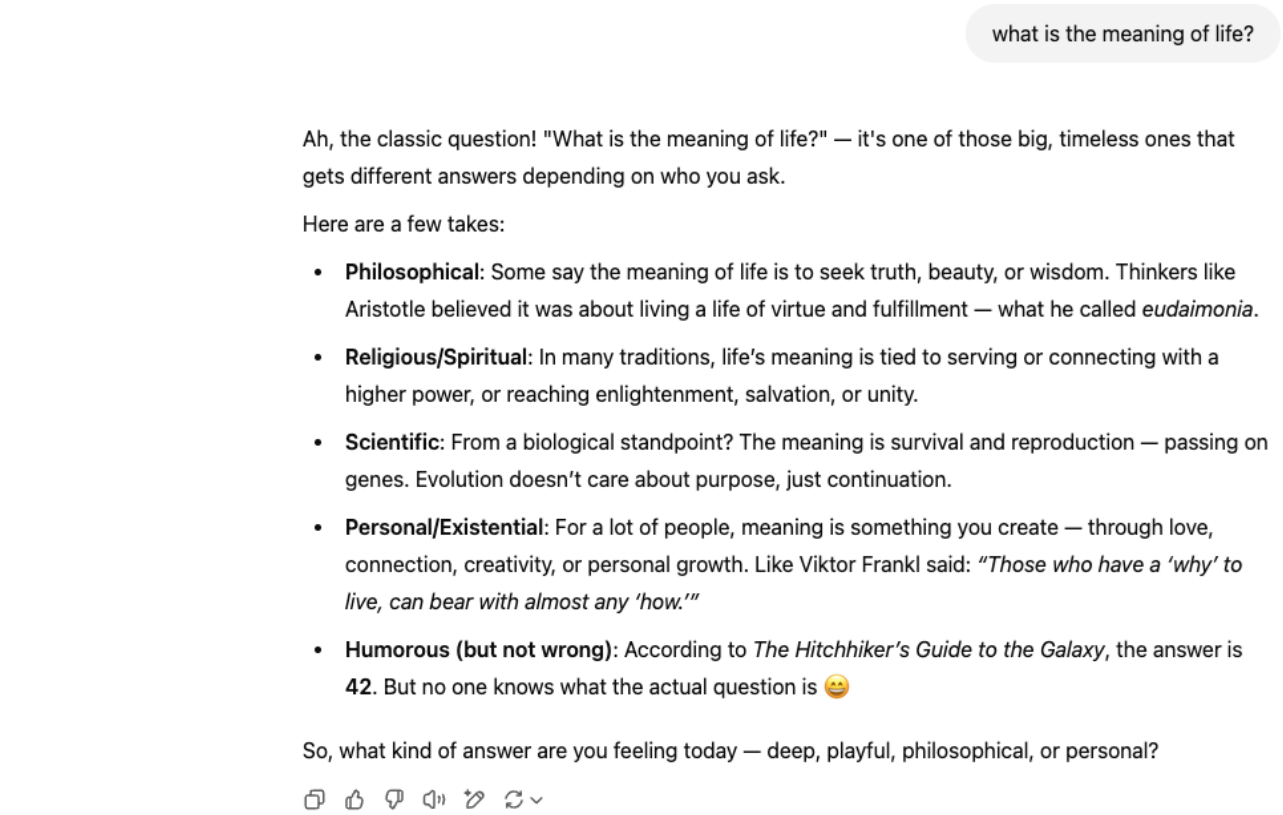
RSAC 2025 Conference

Retrieval Augmented Generation

Why it matters?

Public chatbots

- Supposed to answer all of user's questions


what is the meaning of life?
Ah, the classic question! "What is the meaning of life?" — it's one of those big, timeless ones that gets different answers depending on who you ask.
Here are a few takes:
Philosophical: Some say the meaning of life is to seek truth, beauty, or wisdom. Thinkers like Aristotle believed it was about living a life of virtue and fulfillment — what he called eudaimonia.
Religious/Spiritual: In many traditions, life's meaning is tied to serving or connecting with a higher power, or reaching enlightenment, salvation, or unity.
Scientific: From a biological standpoint? The meaning is survival and reproduction — passing on genes. Evolution doesn't care about purpose, just continuation.
Personal/Existential: For a lot of people, meaning is something you create — through love, connection, creativity, or personal growth. Like Viktor Frankl said: "Those who have a 'why' to live, can bear with almost any 'how.'"
Humorous (but not wrong): According to The Hitchhiker's Guide to the Galaxy, the answer is 42. But no one knows what the actual question is 😄
So, what kind of answer are you feeling today — deep, playful, philosophical, or personal?

But public chatbots don't cut it for the enterprise

How much money did our company make last quarter?

Could you share the specific data or report for last quarter's financials? I'll take a look and help you figure out how much your company made.

LLMs lack

a) Private context
b) Latest information
c) Factual correctness

Adding private data to chatbots

a) Training time: Finetune the model with additional private data

b) Inference time: Provide addition data as context

	In-Context Learning	Fine-tuning
Flexibility	High	Low
Cost	High	Low
Capability	Low	High
Reliability	Low	High

Fine-tuning: better suited for changing model's behavior
In-context: better suited for adding new knowledge

But fine-tuning lost as trends favored in-context learning

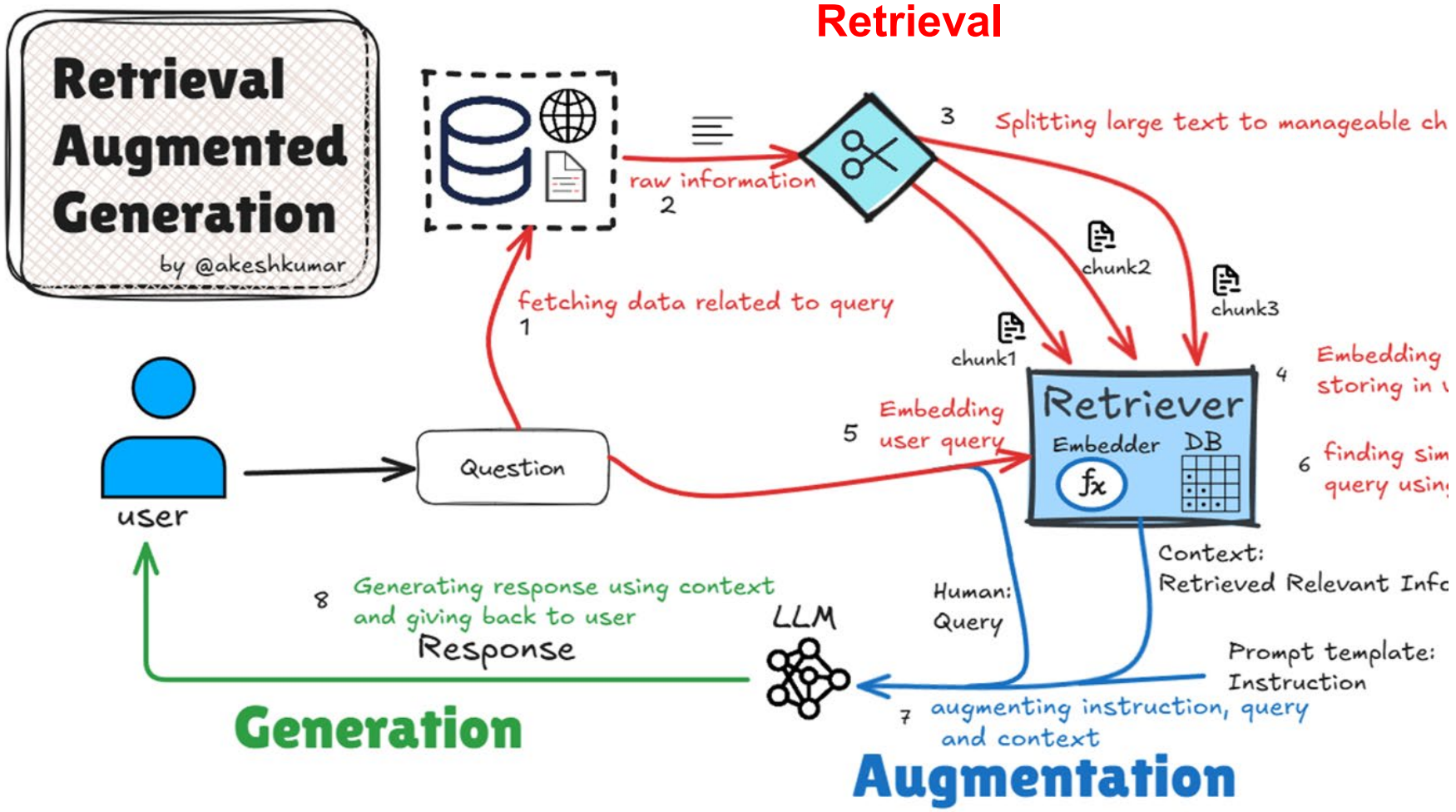
1. Context Window: Context window sizes went from 8K tokens to 1 million tokens

2. Per-token Cost: Cost has dramatically reduced for frontier models

3. Reliability Improvements: LLMs have become better at instruction following and robustness against prompt injections

4. Model size: Frontier models have become incredibly large

RAGs are in-context learning on steroids



RAGs are ideal for enterprise chatbots

- Knowledge base stays outside the model
- Information can be updated by simply updating the knowledge base
- Model can be upgraded without changing the rest
- Factual information can be grounded in real documents, reducing hallucinations.

RSAC 2025 Conference

But security remains a big challenge

Why enterprise RAGs need permissions

- Enterprise data is highly sensitive
- Documents have strict access controls on which user can see what and they must be enforced
- Permissions can frequently (sometimes more frequently than the data itself) change

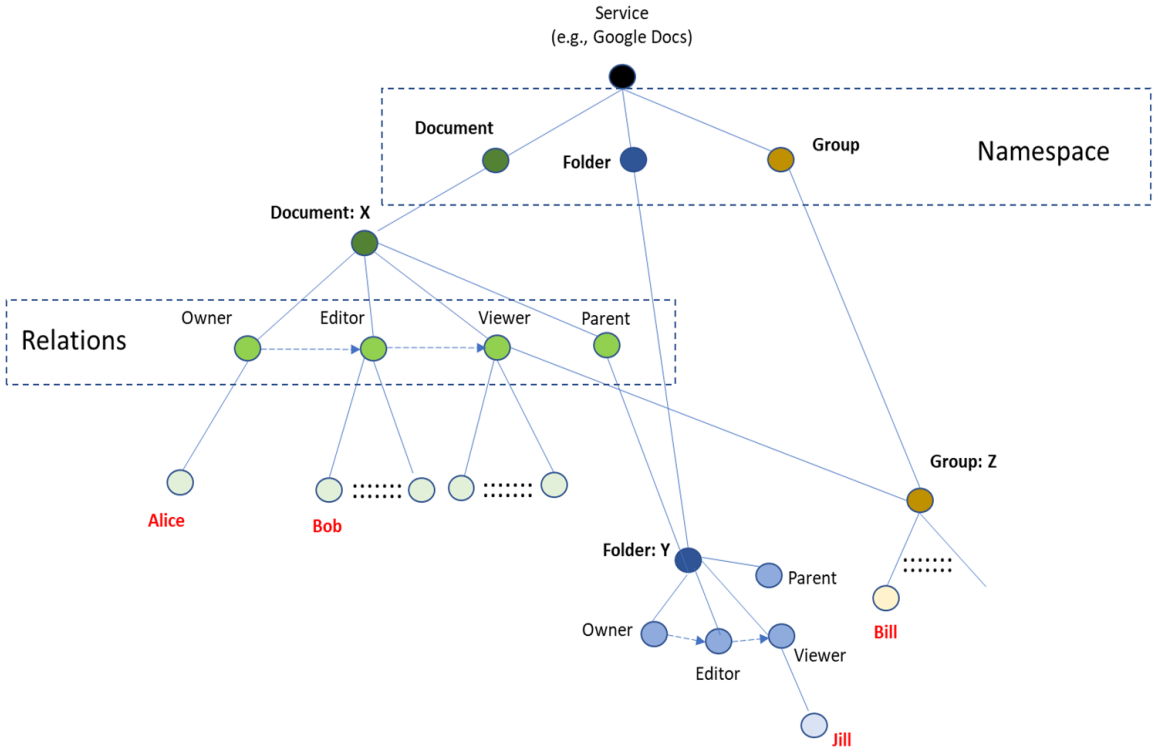
Building a secure knowledge base

Question: Does this user have access to this document?

1. Does the user have a direct access to the document
2. Does the user have access to any folder/directory which may contain this file?
3. Does the user belong to any group or domain that has access to this file?
4. Does the user belong to any group or domain that has access to a folder/directory that may contain this file

Permissions as Graphs

- Permissions complex relationships that evolve and must be adjudicated at run-time



Data Source provides
1. Content
2. File Hierarchy
3. Access Control Lists

Identity Provider provides
1. User Identity
2. User-Group-Domain hierarchy

Enforcing permissions: Pre-filtering

Pre-filtering approach involves first selecting the documents a user has access to and building a knowledge base only on those documents

Advantages:

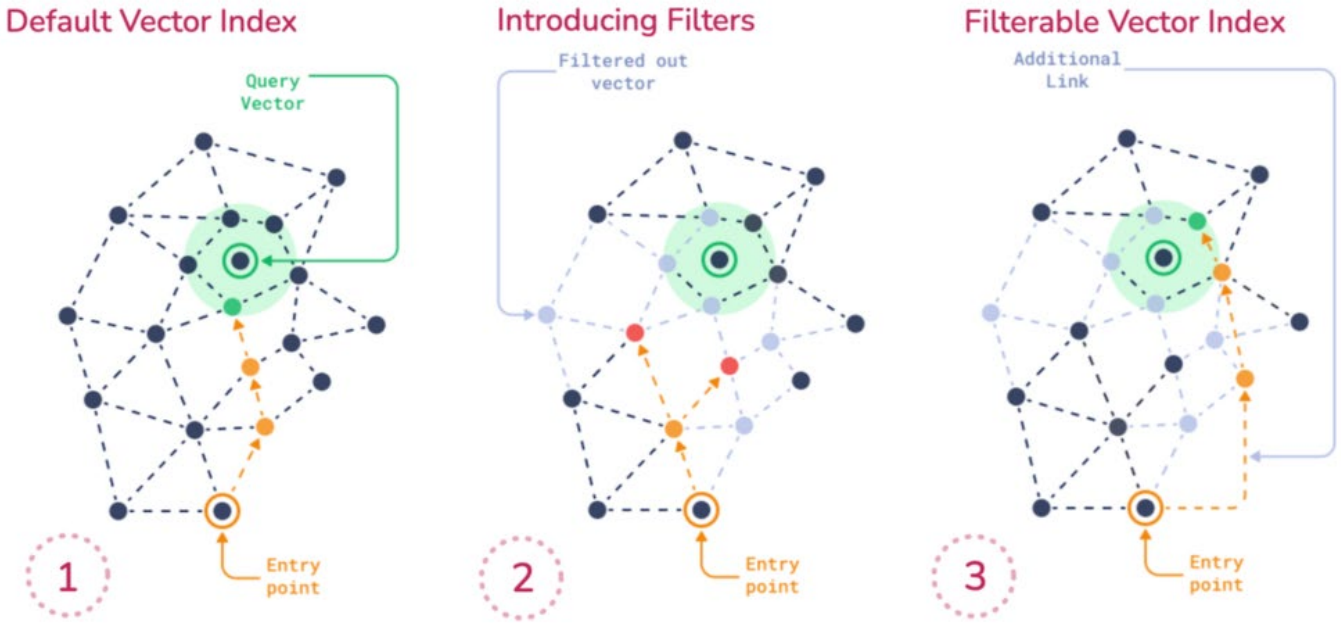
- **Perfect Security**: Data from different users are never mixed
- **Perfect Quality**: The retrieved context is of highest quality

Disadvantages:

- **Low Performance**: Need to make a separate KB for each user
- **Low Flexibility**: Content/Permissions updates are very complex

Enforcing permissions: Metadata filtering

Vector DBs support metadata filtering for labels, time-stamps etc but most of these filters are not suitable for access control



Supported Filters

1. Arithmetic: =<, >=, ==
2. Boolean: AND, OR, NOT
3. Lists: IN, NOT IN
4. Geo-location:
5. Keyword Match: **

These filters are not efficient and suitable for access control

Enforcing permissions: Post-filtering

Post-filtering approach involves building one KB for all documents and then filtering out unauthorized chunks after the fact

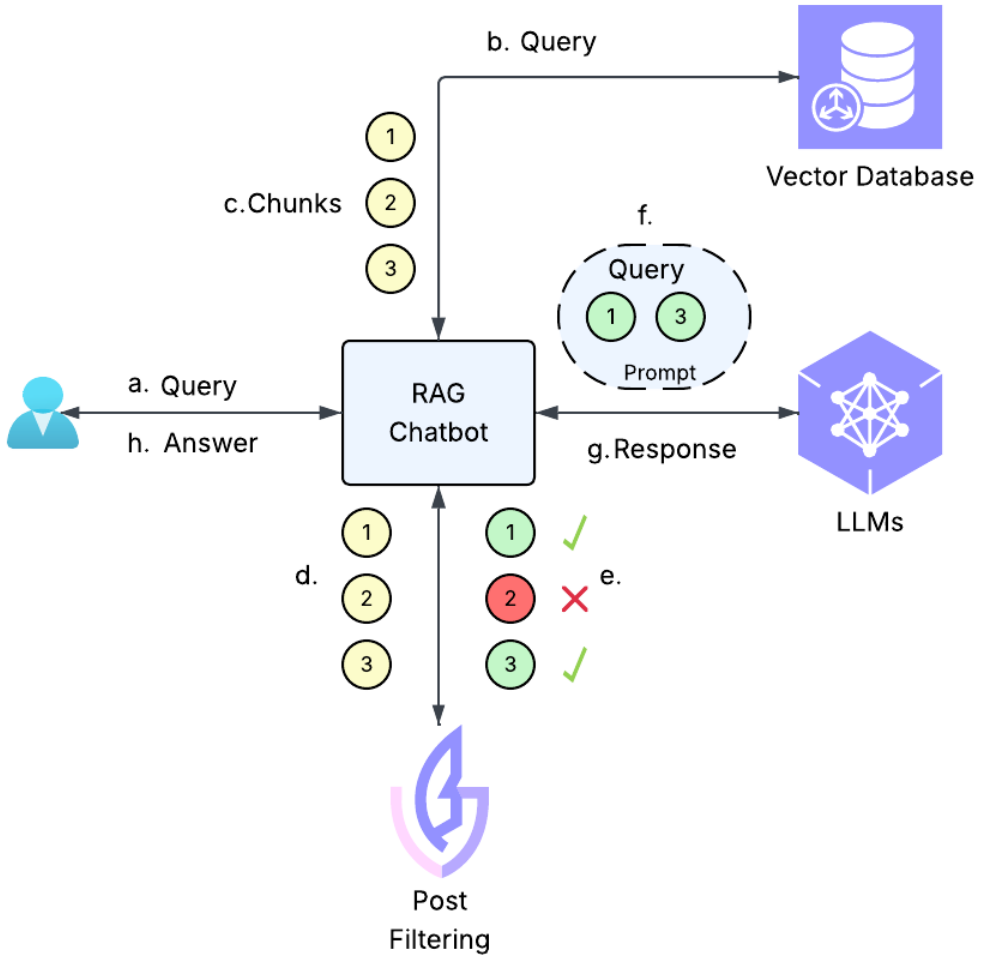
Advantages:

- **Higher Performance**: Only need to build one KB for all users
- **Flexibility**: Content/Permissions can be updated real-time

Disadvantages:

- **Poor security:** Any configuration errors can leak data
- **Lower quality:** Relevant and accessible chunks may be left out of retrieval

A permissioned RAG architecture



RSAC | 2025 Conference

Permissions are not enough

Steps to ensure security of your data

Shared responsibility model

- 👍 Developer is responsible for enforcing permissions as they exist
- 👎 Customer is responsible for ensuring that permissions are correct

Why permissions are not enough

- Permissions are heavily mis-assigned in large organizations.
- 95% of permissions are only used once

POLITICS

The Trump Administration Accidentally Texted Me Its War Plans

U.S. national-security leaders included me in a group chat about upcoming military strikes in Yemen. I didn't think it could be real. Then the bombs started falling.

By Jeffrey Goldberg

If you think that this is not happening in your organization, you are badly mistaken

Excessive access has consequences

STARTUPS
Flexport accuses former employees of stealing its source code to create a rival startup
Charles Rollet — 12:58 PM PDT · March 17, 2025

Marc Benioff
@Benioff

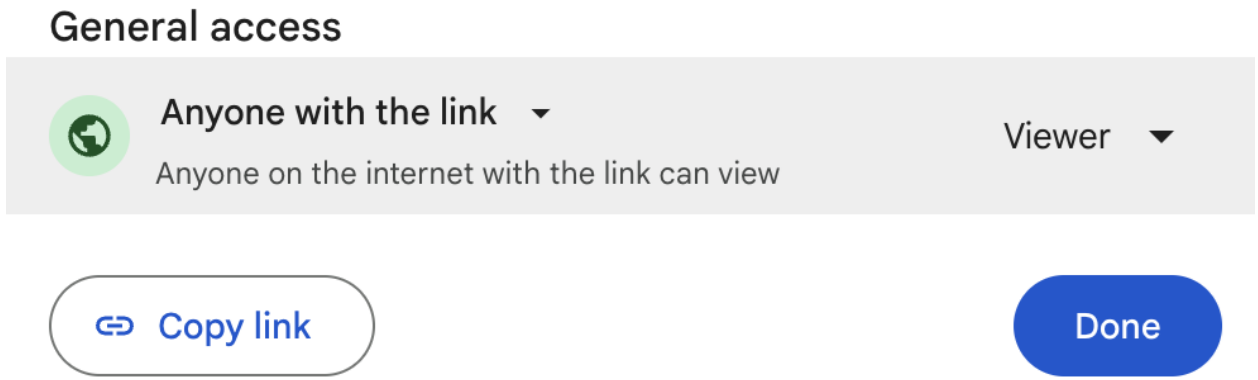
"Microsoft customers deployed Copilot only to discover it can let employees read an inbox or access HR documents. 'Now when Joe Blow logs into an account & kicks off Copilot, they can see everything, All of a sudden Joe Blow can see the CEO's emails.'"

From businessinsider.com

9:13 AM · Nov 21, 2024 · **43.1K** Views

Shared links cause problems

Shared links (documents that can be shared with anyone who has a secret link) are notoriously difficult to track. Their usage in chatbots must be determined by content (M&A should be blocked but Onboarding documents must be allowed)



CoPilot leakage demo

RSAC | 2025 Conference

Going Beyond Permissions

Steps to ensure security of your data

Sensitive data protection goes beyond permissions

1. Sensitive data must be kept out/redacted before reaching the vector index

2. Permissions must not only be enforced but corrected

3. User-AI interactions must be logged and analyzed for breaking terms of use and code of conduct

MIP labels in Microsoft

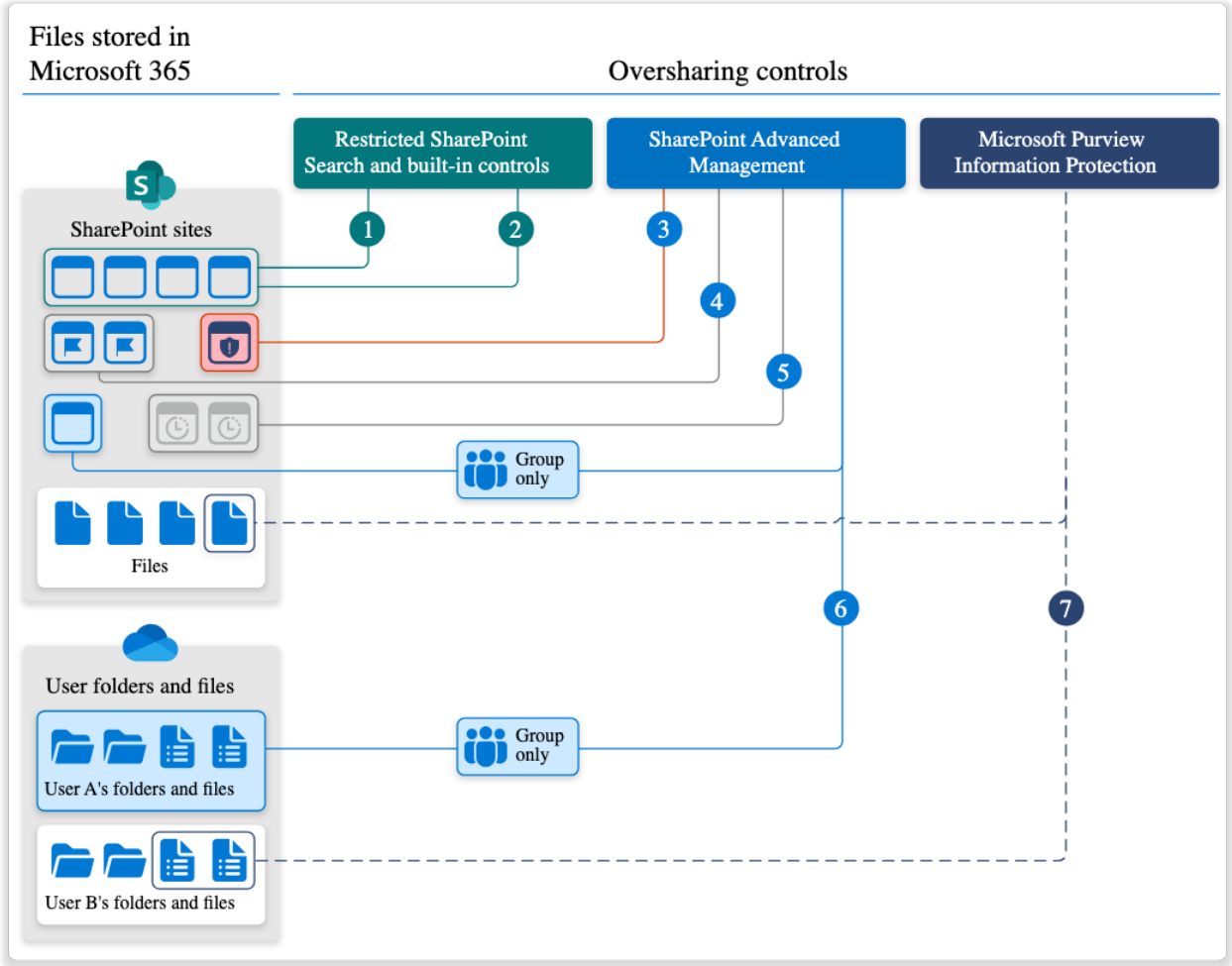
Microsoft Purview Information Protection Labels (MIP) are respected by CoPilot when it comes to indexing or serving content

Problems:

1: Limited to Microsoft ecosystem

2: Poor accuracy especially across different languages and obfuscation techniques

3: No customization for prompts and response analysis



Glean enterprise search

Glean supports a large variety of data sources with their own controls

But,

1. It is permission aware but there is no notion of sensitivity
2. Manual content inclusion and exclusion rules that are very limited

Exclude specific drives/folders

Exclude specific users

Include specific Workspaces

Exclude specific Projects

Clearly this is not enough

1. Permissions are a good starting point but definitely not sufficient
2. We need better security and governance on Enterprise chatbots
3. Shouldn’t pass the buck to the consumer in the name of “shared responsibility model”
4. Must integrate with existing DLP/DSPMs and should not break other user flows
5. Content needs access control, not the files

Apply what we have learned

- Immediate
  - Find out how many knowledge bases have been created
  - If using Copilot or Glean, go through the logs or perform a readiness assessment
- In the short term
  - Catalog your enterprise unstructured data powering AI
- Fix the Security problem
  - Build v/s Buy

Apply what we have learned

- Fixing the Data Security Problem in the AI era requires:
  - Understanding of unstructured data
  - Interoperable system that can broker data flow
  - Resilient to adversarial manipulation
  - Fast to come inline with applications
  - Accurate to avoid causing alert fatigue

RSAC 2025 Conference

Many Voices.
One Community.

Thank You!

Find us on the RSA floor or reach out at help@realmlabs.ai